基于 Ks 分布的被子植物演化的时间尺度研究

焦贝贝[1]， 王希胤[1*]

（1. 华北理工大学 生命科学学院，河北省 唐山 063210）

摘要：物种演化时间估算是生命演化研究的重要部分。近年来，许多研究发现不同基因和不同物种的进化速率差异显著，因此需要新的方法对进化事件发生时间进行重新估计。为了对被子植物演化时间的从新估计，基于共享多倍化事件或共享分歧事件应该有共同 Ks 峰值的理念，建立了基于基因组数据的进化速率矫正模型，结果如下：（1）对获取 Ks 分布三种常见的方式进行了比较分析，明确了通过提取共线性区块上 Ks 值的中位数的方式最优。（2）模拟了 Ks 值随时间累积系数 v 的变化过程下的 Ks 分布。当假设 v 服从正态分布时，Ks 分布出现了长尾现象。（3）将矫正方法在被子植物中的应用，发现不同谱系的被子植物具有同步的辐射进化和适应性进化现象。而且被子植物的进化速率虽然差异显著，但不同分支间的进化速率仍具有部分一致性。例如，木兰类植物进化速率最慢，真双子叶植物次之，单子叶植物进化速率最快。最终得到了相对可靠的物种的演化时间轴，为植物研究提供系统发育和演化的支撑。

关键词：Ks 分布，被子植物，时间矫正，系统发育树，进化速率

中图分类号：Q941 **文献标识码：**A 文章编号：

Time scale of angiosperm evolution based on Ks distribution

JIAO Beibei[1], WANG Xiyin[1*]

（1. *College of Life Sciences*, *North China University of Science and Technology*, Tangshan 063210, Hebei, China）

Abstract: Estimating the time scale of species evolution is an important part of life evolution. It is found that there are significant differences in the evolution rate of different genes and species in recent years, which challenges the molecular clock hypothesis to a great extent. Therefore, new methods are needed to re-estimate the evolutionary events time. The whole genome sequence of angiosperms makes it possible to estimate the evolutionary time from the whole genome perspective. In order to re-estimate the evolution time of angiosperms, an evolution rate correction model based on genomic data is established according to the idea that shared polyploidy events or shared divergence events should have the same Ks peak. The results were as follows: (1) Three common ways to obtain Ks distribution were compared and analyzed, which showed that the best way was to extract the median of Ks values on collinear blocks. (2) The change process of Ks distribution was simulated with time accumulation coefficient v of Ks values. When v was assumed to obey the normal distribution, the Ks distribution had a long tail phenomenon. (3) The correction process was described in detail, which was conducive to the understanding and wide spread of this method. From the application of correction method in angiosperms, it was found that the Ks peak before correction was not linear with time, while the Ks peak after correction was

收稿日期：2022-04-17

基金项目：国家自然科学基金（32070669）[Supported by National Natural Science Foundation of China（32070669）] 。

第一作者：焦贝贝(1992-)，硕士研究生，主要从事比较基因组学研究，（E-mail）jiaobeibei0126@gmail.com。

***通信作者：**王希胤，博士，教授，主要从事比较基因组学研究，（E-mail）wangxiyin@vip.sina.com。

directly proportional to time, indicating that it is very necessary to estimate the time of species evolution events after correcting the Ks peak. It was also found that although the evolution rate of angiosperms was significantly different, the evolution rate between different branches was still partially consistent: Magnoliids had the slowest evolutionary rate, followed by Eudicots and Monocots. When the environment changed greatly, most species of different lineages of angiosperms had synchronous radial evolution and adaptive evolution. Finally, the model corrects angiosperms and established a relatively reliable angiosperm evolution time axis, which helps to understand the evolution process and model, especially to understand the phylogenetic relationship and the causes of diversity and provides phylogenetic and evolutionary support for plant research.

Key words: Ks distribution，angiosperms，time correction，phylogenetic tree，evolutionary rate

被子植物的起源和早期快速演化及其发生时间一直是生物学的研究热点。当前估算物种演化时间的方法主要是基于分子钟假设，即以某几个特定类群的化石时间作为校正点，然后通过部分基因序列间的相似性，假设不同的物种拥有相同或相近的进化速率，来估算系统发育树上某一节点的时间，从而推断出该类群的起源时间（唐先华等，2002；Donoghue & Yang, 2016；Luo et al., 2020）。然而，近年的研究表明，不同物种的分子钟常常具有显著差异，即不同物种的进化速率有显著不同（Wang et al., 2015；Wang et al., 2017；Wang et al., 2018；Wang et al., 2019），不同年代也具有不同的进化速度（罗静和张亚平，2000；Smith & Donoghue, 2008）。而且在不同的研究中，对分子进化速率的估算有很大的差异（Lanfear et al., 2010）。除此之外，引入的化石时间对估算的时间影响很大，随着更多化石且更准确的年份测定，被子植物演化的时间尺度也会随之变动（Hug & Roger, 2007；Wang et al., 2015；Silvestro et al., 2021）。

基因组测序揭示了历史上反复的多倍化事件（Ren et al., 2018），多倍化事件使基因组内所有基因发生重复，且基因组中的古老同源区域常常有相当数目的重复基因保留下来，形成目前基因组内或者基因组间的共线性同源基因（Jiao et al., 2011）。对共线性同源基因的分析，是揭示古代的多倍化或物种分歧事件并推定其发生时间和规模的重要途径。多倍化发生后植物基因组常常会变得很不稳定，进化速率也变得显著不同。由于减少了选择性约束，这些重复基因通常以更快的速度进化（Wang et al., 2016）。例如在葫芦科植物基因组的研究中，发现甜瓜的进化速度最慢，西瓜和黄瓜的进化速度分别快 23.6%和 27.4%（Wang et al., 2018）。

一般认为，同义替换率（synonymous substitution rate，Ks）往往不会改变氨基酸的组成，不受自然选择的影响。因此，Ks 分布常常作为判定物种历史上发生的多倍化或物种分歧事件的依据（Vanneste et al., 2013）。依据共享的演化事件应该有相同的 Ks 峰值，Wang 等人首次提出了基于 Ks 峰值的矫正方法用以估算物种演化的时间尺度，得到了其他研究工作者的认可，还被广泛应用于他们的研究中（Zhuang et al., 2019；Shang et al., 2020；Song et al., 2020；Yang et al., 2020；Song et al., 2021；Wang et al., 2021）。例如两个团队分别对睡莲（Zhang et al., 2020a）和芡实（Yang et al., 2020）基因组分析，Yang 等人通过 Ks 峰值矫正的方式估算的芡实古老多倍化（被证实为睡莲目共享）与另外的团队基于睡莲目的转录组数据估算的时间尺度基本一致。基于 Ks 峰值的矫正方法中，获得准确的 Ks 峰是准确估算时间尺度的关键。然而，当前获取 Ks 分布的方式不统一而且常常带有长尾现象（Tang et al., 2008）。为何 Ks 分布会有长尾现象？长尾现象对 Ks 峰是否有重要影响等问题，也尚未有清晰的表述。

目前，已有 400 余种被子植物的基因组得到不同水平的测定，便于在全基因组的尺度上理解这些被子植物的演化历程（Kress et al., 2022）。全基因组数据能有效消除横向基因转

移和类群间基因进化速率差异等因素对系统发育树的影响。因此，急需在全基因组数据层面上，利用新的方法对被子植物的演化时间进行重新估计。该文拟对三种获取 Ks 分布的方式进行比较，明确哪种方式获得的 Ks 峰值更接近真实情况；对于 Ks 分布中常见的长尾现象，采用模拟仿真的方式，探究出现长尾现象的原因；区分共享多倍化和共享早期分化两种情况，创建基于全基因组数据的 Ks 分布矫正模型，对 44 个代表性被子植物基因组演化事件的时间尺度进行重新估计，得到相对可靠的被子植物演化时间轴。这有助于更深层地了解被子植物多样性和系统发育以及被子植物基因组的进化模式。

1 材料与方法

1.1 基因组数据材料

收集 44 个高质量染色体水平的被子植物基因组（主要来自 NCBI 和 PHYTOZOME），共包含 43 科 39 目（表 1）。

1.2 方法

1.2.1 共线性分析

使用 WGDI v0.5.3（Sun et al., 2021）软件进行共线性分析。首先，使用 BLASTP 来识别基因组内或基因组间的基因相似性。随后，用 WGDI 软件的‘-d’子程序绘制同源点阵图，并运行‘-icl’子程序获得共线性基因。

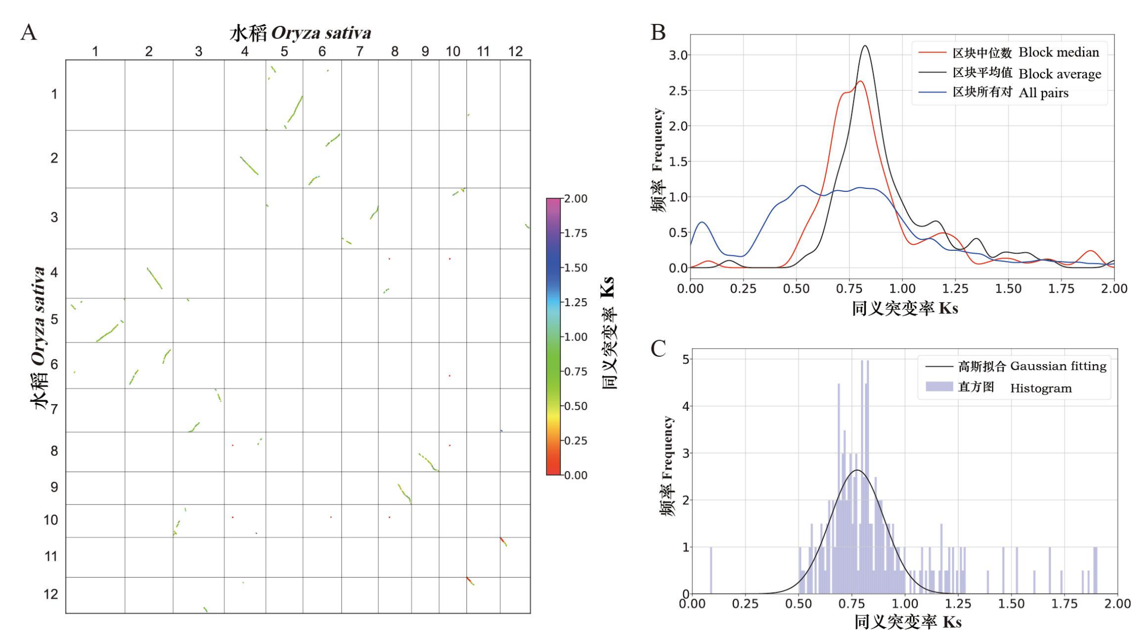
1.2.2 Ks 分布

Ks 分布主要是通过 WGDI 软件完成的。首先，使用 WGDI 软件的‘-ks’子程序调用 PAML（Yang, 2007）软件计算共线性基因对的 Ks 值。通过‘-bi’子程序整合共线性和 Ks 值的结果，并使用 WGDI 软件的‘-bk’子程序查看共线基因的 Ks 值的分布，结果以点图的形式展示（图 1：A）。根据物种内或种间已知的多倍化或分歧事件，通过 WGDI 的‘-c’子程序对共线性片段进行过滤，只保留多倍化事件或分歧事件产生的共线性片段。然后，通过 WGDI 的‘-kp’子程序获取 Ks 分布（图 1：B）。最后，使用 WGDI 中的“-pf”子程序对不同事件分别进行拟合并获取 Ks 分布（图 1：C）。

2 结果与分析

2. 1 Ks 分布和长尾现象解析

Ks 分布常常用来判定物种历史上发生的多倍化或物种分歧事件的依据。目前获取 Ks 分布主要有三种方式。方式一：通过 OrthoMCL（Li et al., 2003）等聚类软件获取旁系同源基因对，然后计算这些同源基因对的 Ks 值并绘制 Ks 分布图。方式二：先进行基因组共线性分析，然后计算共线性基因对 Ks 值并绘制 Ks 分布。方式三：在方式二的基础上，提取共线性区块上 Ks 值的中位数并绘制 Ks 分布。三种方式中，方式一由于没有共线性分析，所获取的旁系同源基因对常常会有大量串联重复基因从而影响 Ks 分布。方式二和三都经过了共线性分析，当把共线性区块（长度大于 5）上同源基因对的 Ks 值以点图的形式展示出来时（图 1：A），这里以水稻为例，可以看到大部分由绿色的点组成的片段，例如 8 号与 9 号染色体，这与水稻近期的一次多倍化事件相符。Ks 点图中大部分点的颜色相近，说明 Ks 值波动很小。对共线性区域的 Ks 值的中位数（方式三）、平均值和所有的基因对（方式二）进行正态分布拟合（带宽为 0.01，homo 范围 0.3-1）（图 2：B），可以看到方式二并没有产生明显的峰，而且 Ks 分布整体带有长长的尾巴。方式三和区块的平均值的 Ks 分布有明显峰值，数据更为集中。由于中位数是对总体中心很好的估计，且稳健性更强，中位数的峰值颜色和 Ks 点图的颜色更为接近，因此区块的 Ks 值的中位数更接近 Ks 真正的峰值，对方式三的 Ks 分布按照正态分布拟合来提取 Ks 峰值（图 1：C）。



A. 水稻基因组的共线性区块；**B.** 共线性区块上 Ks 值的拟合分布；**C.** 共线性区块 Ks 值的核密度估计。

A. Synteny blocks of the *Oryza sativa* genome；**B.** Fitted distribution of Ks values for synteny blocks；**C.** Kernel density of Ks values for synteny blocks.

图 1　Ks 分布

Fig.1　Ks distribution

为了进一步解析长尾现象，模拟了 Ks 分布随进化速率的演变过程。假设最初的 Ks 分布服从正态分布 $X \sim N(\mu,\sigma^2)$，其中期望 μ（峰值）和标准差 σ 为常数。分子钟理论认为基因的进化速率是相对恒定的，因此定义 v $(v>1)$，代表 Ks 值的时间累积系数，表示初始 Ks 值随时间演化不断累积，模拟真实情况下的恒定进化速率。而其他研究表明分子钟也并非等速进行，同时假设 v 也服从正态分布 $X_v \sim N(\mu_v,\sigma_v{}^2)$，对这两种假设分别进行了数据仿真模拟。Ks 值随着时间的推移进行迭代为 X'，迭代次数为 n。当 v 为常数值时，

$$X' = X * v^n$$

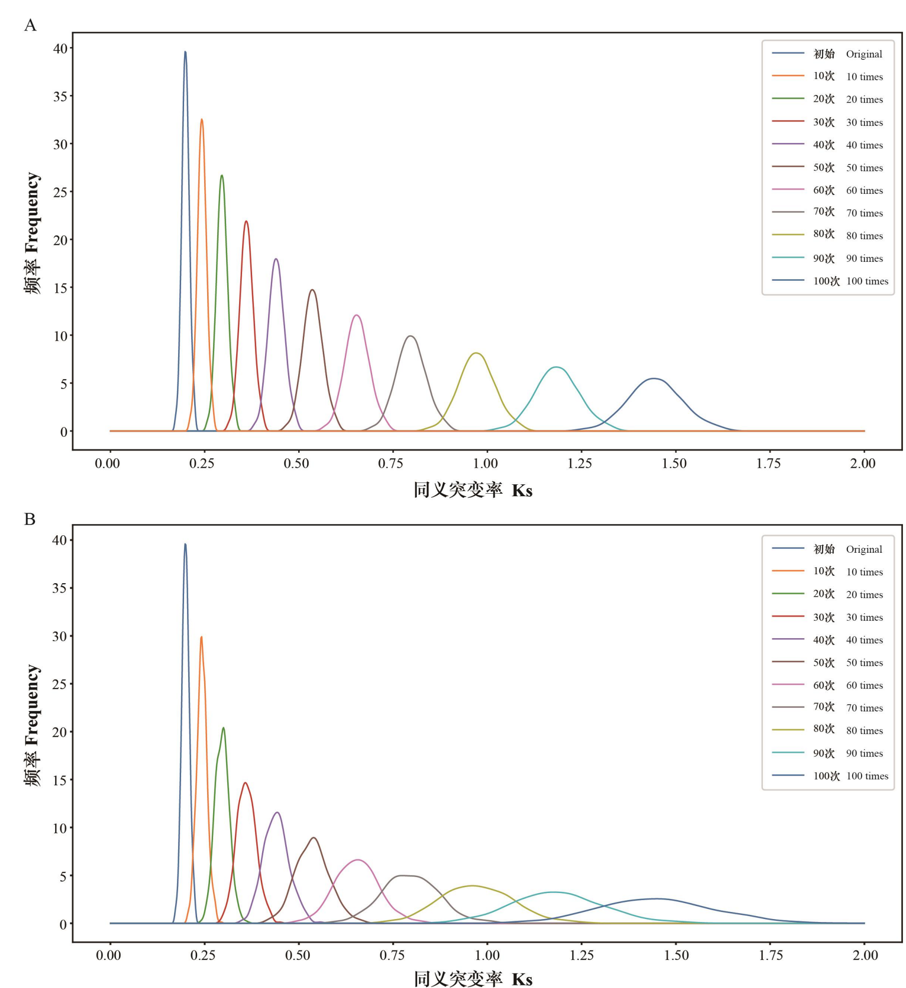
当 v 服从正态分布时，

$$X' = X * X_v{}^n$$

当假设 Ks 值的时间累积系数 v 为一个常数值时，设置假设的 Ks 分布为 $X \sim N(\mu,\sigma^2)$，依据 Ks 分布数据特征，设定 $\mu = 0.2$，$\sigma = 0.01$，$v = 1.02$，$n = 100$。每迭代 10 次，绘制 Ks 分布结果（图 2：A）。随着进化事件的推移，Ks 峰值也逐渐变大，Ks 分布依旧完美符合正态分布且没有明显的长尾分布现象。

当假设 Ks 值的时间累积系数 v 服从正态分布时，最初设置假设的 Ks 分布为 $X_v \sim N(\mu_v,\sigma_v{}^2)$，其中 $\mu = 0.2$，$\sigma = 0.01$，$\mu_v = 1.02$，$\sigma_v = 0.01$，$n = 100$。每迭代

10 次，绘制 Ks 分布结果（图 2：B）。随着进化事件的推移，Ks 峰值也逐渐变大，Ks 分布也不再是正态分布，并带有明显的长尾现象。这种假设所得到的结果更接近于真实情况，因此，基因的进化速率不是相对恒定的，它可能并非等速进行，而是在不同年代具有不同的进化速度，这可能也符合正态分布。对模拟的 Ks 分布再次通过高斯拟合获取峰值时，发现 Ks 峰值与进化速率匀速时的没有明显差异（表 2）。因此，Ks 分布中长尾现象对提取到的 Ks 峰值的影响较小。



A. Ks 分布在恒定进化速率下的模拟；**B.** Ks 分布在进化速率服从正态分布的模拟。

A. Simulation of Ks distribution at a constant evolution rate；**B.** Simulation of Ks distribution under a normal distribution of evolution rates.

图 2　Ks 分布在不同进化速率下的模拟结果

Fig.2　Simulation results of Ks distribution at different evolution rates

表 2　不同进化速率模拟下的 Ks 峰值

Table 2　Ks peaks under simulations at different evolution rates

迭代次数 Number of iterations	匀速分布 Uniform distribution	正态分布 Normal distribution	差异 Difference
0	0.200	0.200	0.000
10	0.244	0.243	-0.001
20	0.297	0.297	0.000
30	0.362	0.355	-0.007

40	0.442	0.431	0.010
50	0.538	0.531	0.008
60	0.656	0.650	-0.006
70	0.800	0.792	-0.008
80	0.975	0.961	-0.014
90	1.189	1.180	-0.009
100	1.449	1.442	-0.007

2.2 Ks 分布矫正方法

被子植物基因组常常经历不止一次多倍化事件，由于不同物种的进化速率显著不同，就导致共享的多倍化事件的 Ks 峰值也大不相同。而 Ks 分布矫正方法的核心理念就是将这些共享事件的 Ks 峰矫正到一起。根据共享事件的不同，Ks 分布矫正方法又分为共享多倍化和共享分化两种情况。

如果物种 A、B 存在共享的多倍化事件，那么这次多倍化事件在不同物种中发生的时间应该是相同的，Ks 峰值也应该是相等的（图 3：A）。黄色方块代表两个物种共享的多倍化，即 $Ks_{AA}=Ks_{BB}$，对应的时间范围为物种 A、B 从多倍化事件到当前的时间点（绿色的大括号）。由于物种不同的进化速率，现实情况下的 Ks_{AA} 和 Ks_{BB} 并不相等。假设多倍化事件之后物种 A 和 B 有各自的进化速率分别为 v_A 和 v_B，O 是物种 A、B 的分化节点，从多倍化事件到分歧点 O，物种 A、B 的祖先拥有的进化速率为 v。那么，物种 A 的进化速率 v_A 要想恢复到 v 就要乘以它的矫正系数为 $\lambda_A=\frac{v}{v_A}$。同理，物种 B 的矫正系数为 $\lambda_B=\frac{v}{v_B}$。因而，物种 A、B 间分化的 Ks_{AB} 矫正后为 $Ks_{AB-correction}=Ks_{AB}\lambda_A\lambda_B$（Yang et al., 2020）。

如果两个物种 A、B 不存在共享的多倍化事件但存在共享的早期分化事件时，通过寻找外类群来辅助矫正（图 3：B）。物种 C、D、E 是外类群，物种 C 和 D 的祖先在 P 点与物种 A、B 的祖先分化，所以物种 C 与 A、B 间的 Ks 峰值应该相等，物种 D 与 A、B 间的 Ks 峰值也应该相等，即 $Ks_{CA}=Ks_{CB}$，$Ks_{DA}=Ks_{DB}$。同样，由于物种间不同的进化速率，现实情况下它们大多是不相等的。按照前面的假设，

$$\frac{Ks_{CA\text{-}correction}}{Ks_{CB\text{-}correction}}=\frac{Ks_{CA}\lambda_C\lambda_A}{Ks_{CB}\lambda_C\lambda_B}=\frac{Ks_{CA}\lambda_A}{Ks_{CB}\lambda_B}=1,\quad 即\frac{\lambda_A}{\lambda_B}=\frac{Ks_{CB}}{Ks_{CA}}。$$

同理，

$$\frac{Ks_{DA\text{-}correction}}{Ks_{DB\text{-}correction}}=\frac{Ks_{DA}\lambda_D\lambda_A}{Ks_{DB}\lambda_D\lambda_B}=\frac{Ks_{DA}\lambda_A}{Ks_{DB}\lambda_B}=1,\quad 即\frac{\lambda_A}{\lambda_B}=\frac{Ks_{DB}}{Ks_{DA}}。$$

可以发现，当选取的外类群越多，获取的 λ_A 和 λ_B 的关系越准确。最后取平均值表示它们之

间的关系 $\frac{\lambda_A}{\lambda_B}=mean(\frac{Ks_{CB}}{Ks_{CA}},\frac{Ks_{DB}}{Ks_{DA}},...)$。

A. 共享多倍化事件；**B.** 共享早期分化。

A. Shared polyploidy events；**B.** Shared early divergence.

图 3　Ks 分布矫正方法的原理

Fig.3　Principle of the Ks distribution correction method

2.3 被子植物系统发育树时间矫正

目前，很多用系统发育树的方法推测被子植物的演化时间，认为被子植物的起源为三叠纪 225 百万年至 240 百万年前（Magallón, 2010），这与起传粉作用的核心植食性鳞翅目昆虫的起源时间（约 230 百万年前）一致（Li et al., 2019）。由于无油樟目和睡莲目、核心被子植物五大分支之间的关系仍然没有完全解析，已有多个证据暗示核心被子植物祖先可能发生了快速辐射分化（Yang et al., 2020）。因此，在矫正过程中，以无油樟目为作为参考，不讨论它和睡莲目的关系，并且认为五大分支的分化时间尺度在同一个时间范围内。基于核心真双子叶植物共享的 γ 事件，时间范围为 115-130 百万年（Million years ago，Mya），对 44 个被子植物基因组（表 1）进行了时间尺度矫正（图 4）。从矫正后的时间尺度来看，被子植物在 130 百万年前附近，单子叶植物、真双子叶植物、木兰类植物祖先都发生了快速辐射进化，与先前的结论一致（Zhang et al., 2020b）。此外，在早白垩世（130 百万年）时期，白垩纪-古新世（K-Pg）边界时期（66 百万年）和中新世（20 百万年，靠近冰川期）很多被子植物发生了的多倍化事件，与先前的研究发现 WGD 的时间在被子植物的系统发育中并不是随机分布的结论一致（Wu et al., 2020）。

尽管不同物种的进化速率数值显著不同，但是同一类群中的进化速率往往具有部分一致性。由矫正方法可知，矫正后的 Ks 峰应该相等。因此，Ks 峰值越大，表明进化速率越快。对木兰类植物、真双子叶植物和单子叶植物与无油樟的 Ks 峰值的比较发现，木兰类植物（大多数为木本）进化速率最慢，真双子叶植物（大多数为灌木）次之，单子叶植物（大多数草本）进化速率最快（表 3），这与多年生木本植物比草本植物的分子进化速率快的结论相符（Lanfear et al., 2013）。此外，对多倍化事件发生的时间与矫正前后的 Ks 峰值比较发现（图 5），矫正前的 Ks 峰值与时间并不是线性关系。随着 Ks 峰值的增大，多倍化事件发生的时间并没有更古老。而矫正后的 Ks 峰值与时间成正比，因此，对 Ks 峰值进行矫正之后再估算物种演化事件的时间是十分必要的。

图 4　时间矫正后的被子植物系统发育树

Fig.4　Angiosperm phylogenetic tree after time correction

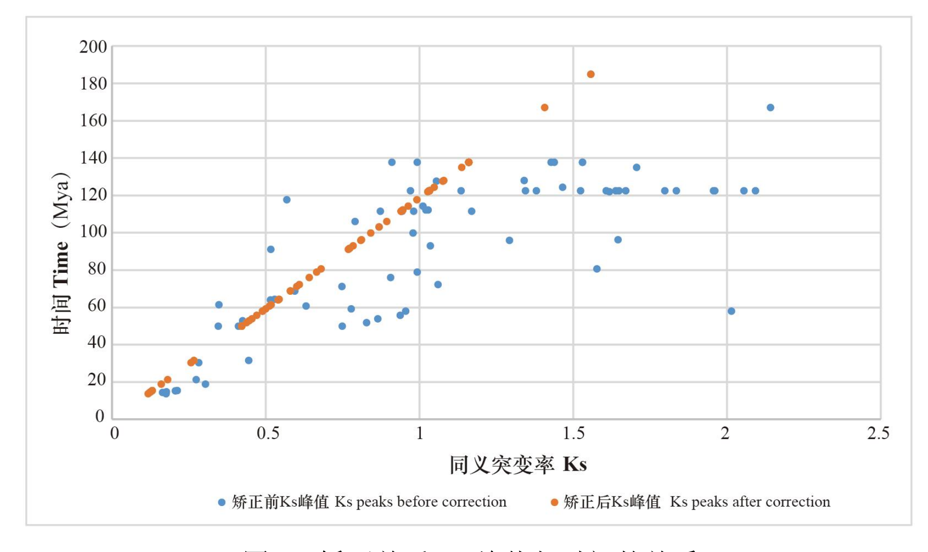


图 5　矫正前后 Ks 峰值与时间的关系

Fig.5　Relationship between Ks peaks and time before and after correction

表 1 研究所用的 44 个被子植物及基因组数据来源

Table 1 List of the 44 angiosperms involved and the genome data sources

物种 Species	目 Order	科 Family	数据来源 Data source
无油樟 *Amborella trichopoda*	无油樟目 Amborellales	无油樟科 Amborellaceae	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/471/905/GCF_000471905.2_AMTR1.0/
蓝星睡莲 *Nymphaea colorata*	睡莲目 Nymphaeales	睡莲科 Nymphaeaceae	https://phytozome-next.jgi.doe.gov/info/Ncolorata_v1_2
鹅掌楸 *Liriodendron chinense*	木兰目 Magnoliales	木兰科 Magnoliaceae	https://www.ncbi.nlm.nih.gov/genome/45466
牛樟 *Cinnamomum kanehirae*	樟目 Laurales	樟科 Lauraceae	https://www.ncbi.nlm.nih.gov/genome/57158
卷毛马兜铃 *Aristolochia fimbriata*	胡椒目 Piperales	马兜铃科 Aristolochiaceae	https://ngdc.cncb.ac.cn/search/?dbId=gwh&q=Aristolochia&page=1
柳叶蜡梅 *Chimonanthus salicifolius*	樟目 Laurales	蜡梅科 Calycanthaceae	https://www.ncbi.nlm.nih.gov/genome/82066?genome_assembly_id=1651656
紫萍 *Spirodela polyrhiza*	泽泻目 Alismatales	天南星科 Araceae	https://data.jgi.doe.gov/refine-download/phytozome?organism=Spolyrhiza&expanded=290
水稻 *Oryza sativa*	禾本目 Poales	禾本科 Poaceae	https://phytozome-next.jgi.doe.gov/info/Osativa_v7_0
菠萝 *Ananas comosus*	禾本目 Poales	凤梨科 Bromeliaceae	https://phytozome-next.jgi.doe.gov/info/Acomosus_v3
椰子 *Cocos nucifera*	棕榈目 Arecales	棕榈科 Arecaceae	The genome draft of coconut (Cocos nucifera)
油棕 *Elaeis guineensis*	棕榈目 Arecales	棕榈科 Arecaceae	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/442/705/
小果野蕉 *Musa acuminata*	姜目 Zingiberales	芭蕉科 Musaceae	https://www.ncbi.nlm.nih.gov/genome/?term=HE813975%E2%80%93HE813985
鼓槌石斛 *Dendrobium chrysotoxum*	天门冬目 Asparagales	兰科 Orchidaceae	https://www.ncbi.nlm.nih.gov/genome/41833
文竹 *Asparagus setaceus*	天门冬目 Asparagales	天门冬科 Asparagaceae	https://datadryad.org/stash/dataset/doi:10.5061/dryad.1c59zw3rm
金鱼藻 *Ceratophyllum demersum*	金鱼藻目 Ceratophyllales	金鱼藻科 Ceratophyllaceae	https://genomevolution.org/CoGe/GenomeInfo.pl?gid=56569

莲 *Nelumbo nucifera*	山龙眼目 Proteales	莲科 Nelumbonaceae	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/365/185/GCF_000365185.1_Chinese_Lotus_1.
昆栏树 *Trochodendron aralioides*	昆栏树目 Trochodendrales	昆栏树科 Trochodendraceae	http://gigadb.org/dataset/view/id/100657/File_page/5
洛杉矶耧斗菜 *Aquilegia coerulea*	毛茛目 Ranunculales	毛茛科 Ranunculaceae	https://data.jgi.doe.gov/refine-download/phytozome?organism=Acoerulea&expanded=322
油蜡树 *Simmondsia chinensis*	石竹目 Caryophyllales	油蜡树科 Simmondsiaceae	https://ngdc.cncb.ac.cn/search/?dbId=gwh&q=GWHAASQ00000000
中华猕猴桃 *Actinidia chinensis*	杜鹃花目 Ericales	猕猴桃科 Actinidiaceae	ftp://www.whiteflygenomics.org/pub/kiwifruit/A_chinensis/Red5/v1.0/Red5_genome_v1.0.fa.gz
杜鹃 *Rhododendron simsii*	杜鹃花目 Ericales	杜鹃花科 Ericaceae	https://www.ncbi.nlm.nih.gov/genome?LinkName=bioproject_genome&from_uid=588298
油橄榄 *Olea europaea*	唇形目 Lamiales	木樨科 Oleaceae	https://www.ncbi.nlm.nih.gov/genome/?term=Olea+europaea
大星牵牛 *Ipomoea trifida*	茄目 Solanales	旋花科 Convolvulaceae	https://datadryad.org/stash/dataset/doi:10.5061/dryad.b9m61cg
中粒咖啡 *Coffea canephora*	龙胆目 Gentianales	茜草科 Rubiaceae	https://www.ncbi.nlm.nih.gov/genome/?term=Coffea+canephora
红花 *Carthamus tinctorius*	菊目 Asterales	菊科 Asteraceae	https://safflower.scuec.edu.cn/download.html
芹菜 *Apium graveolens*	伞形目 Apiales	伞形科 Apiaceae	http://celerydb.bio2db.com
葡萄 *Vitis vinifera*	葡萄目 Vitales	葡萄科 Vitaceae	http://www.grapegenomics.com/pages/VvCabSauv/download.php
野黄瓜 *Cucumis hystrix*	葫芦目 Cucurbitales	葫芦科 Cucurbitaceae	https://figshare.com/articles/dataset/Genome_assembly_of_Cucumis_hystrix/13377671
菜豆 *Phaseolus vulgaris*	豆目 Fabales	豆科 Fabaceae	https://www.ncbi.nlm.nih.gov/genome/?term=Phaseolus+vulgaris
雷公藤 *Tripterygium wilfordii*	卫矛目 Celastrales	卫矛科 Celastraceae	https://www.ncbi.nlm.nih.gov/genome/12874
苹果 *Malus domestica*	蔷薇目 Rosales	蔷薇科 Rosaceae	https://www.ncbi.nlm.nih.gov/genome/?term=Malus+domestica

欧洲大叶杨 *Populus trichocarpa*	金虎尾目 Malpighiales	杨柳科 Salicaceae	https://www.ncbi.nlm.nih.gov/genome/?term=Populus+trichocarpa
垂枝桦 *Betula pendula*	壳斗目 Fagales	桦木科 Betulaceae	https://genomevolution.org/CoGe/GenomeInfo.pl?gid=35080
杨桃 *Averrhoa carambola*	酢浆草目 Oxalidales	酢浆草科 Oxalidaceae	https://ngdc.cncb.ac.cn/search/?dbId=gwh&q=GWHABKE00000000
可可树 *Theobroma cacao*	锦葵目 Malvales	锦葵科 Malvaceae	https://www.ncbi.nlm.nih.gov/genome/?term=Theobroma+cacao
大桉 *Eucalyptus grandis*	桃金娘目 Myrtales	桃金娘科 Myrtaceae	https://www.ncbi.nlm.nih.gov/genome/?term=Eucalyptus+grandis
漾濞槭 *Acer yangbiense*	无患子目 Sapindales	无患子科 Sapindaceae	http://gigadb.org/dataset/100610
珙桐 *Davidia involucrata*	山茱萸目 Cornales	蓝果树科 Nyssaceae	https://ngdc.cncb.ac.cn/search/?dbId=gwh&q=%20PRJCA001721&page=1
伯乐树 *Bretschneidera sinensis*	十字花目 Brassicales	叠珠树科 Akaniaceae	https://www.ncbi.nlm.nih.gov/genome/?term=GCA_018105755.1
连香树 *Cercidiphyllum japonicum*	虎耳草目 Saxifragales	连香树科 Cercidiphyllaceae	https://doi.org/10.1111/nph.16798
四川金粟兰 *Chloranthus sessilifolius*	金粟兰目 Chloranthales	金粟兰科 Chloranthaceae	https://github.com/yongzhiyang2012/Chloranthus-sessilifolius-genome/tree/main/Annotation
参薯 *Dioscorea alata*	薯蓣目 Dioscoreales	薯蓣科 Dioscoreaceae	https://phytozome-next.jgi.doe.gov/info/Dalata_v2_1
芒苞草 *Acanthochlamys bracteata*	露兜树目 Pandanales	翡若翠科 Velloziaceae	https://www.ncbi.nlm.nih.gov/genome/?term=PRJNA703828
滇南黄杨 *Buxus austroyunnanensis*	黄杨目 Buxales	黄杨科 Buxaceae	https://datadryad.org/stash/dataset/doi:10.5061/dryad.cjsxksn6d

表 3 部分核心被子植物与无油樟之间的 Ks 峰值

Table 3 Ks peaks between some species of Mesangiospermae and *Amborella trichopoda*

核心被子植物 Mesangiospermae	物种 Species	峰值 Ks peaks	平均值 Average values
真双子叶植物 Eudicots	昆栏树 *Trochodendron aralioides*	1.671	1.743
	连香树 *Cercidiphyllum japonicum*	1.700	

	滇南黄杨 *Buxus austroyunnanensis*	1.750	
	葡萄 *Vitis vinifera*	1.804	
	洛杉矶耧斗菜 *Aquilegia coerulea*	1.789	
木兰类植物 Magnoliids	鹅掌楸 *Liriodendron chinense*	1.628	1.651
	牛樟 *Cinnamomum kanehirae*	1.642	
	柳叶蜡梅 *Chimonanthus salicifolius*	1.684	
单子叶植物 Monocots	水稻 *Oryza sativa*	2.273	2.082
	紫萍 *Spirodela polyrhiza*	2.210	
	椰子 *Cocos nucifera*	1.834	
	菠萝 *Ananas comosus*	2.106	
	芒苞草 *Acanthochlamys bracteata*	2.117	
	参薯 *Dioscorea alata*	1.950	

3 讨论与结论

长期以来，估算被子植物演化的时间尺度主要是基于分子钟假设，然而分子进化异速现象的广泛存在严重影响其准确性，Wang 等人提出的基于 Ks 分布的矫正方法，获得了令人信服的时间尺度。该文对获取 Ks 分布三种常见的方式进行了比较分析，明确了通过提取共线性区块上 Ks 值的中位数更能代表真实的 Ks 峰值。还进一步解析 Ks 分布中常见的长尾现象，模拟结果表明基因的进化速率不是相对恒定的，它并非等速进行。当假设进化速率不是相对恒定的，而是符合正态分布的时候，Ks 分布出现了有明显的长尾现象，但这并不影响提取到的 Ks 峰值的准确性。先前的研究表明，当 Ks 值大于 1 时，容易受到饱和效应的影响，而且随着 Ks 值增大，这种效应越明显（Vanneste et al., 2013）。模拟的 Ks 峰值范围接近于 1，随着 Ks 峰值增大，估计的 Ks 峰值可能会受到饱和效应的影响。

本研究还详细描述了基于 Ks 峰值的矫正方法的矫正过程。先前的研究只对共享多倍化和共享早期分化两种情况分开进行了描述，这是首次全面的描述，有助于深入理解和传播。基于该方法，还对 44 个高质量的被子植物基因组演化事件的时间尺度进行了重新估计，估计结果与近期发表的时间尺度基本一致（Li et al., 2019；Wu et al., 2020）。结果还表明被子植物基因组的进化速率差异显著，但不同分支间的进化速率仍具有一致性。而且不同谱系的被子植物具有同步的辐射进化和适应性进化现象。随着更多高质量的被子植物基因组的公布和有效化石年份的准确测定，被子植物演化的时间尺度会越来越清晰，更有利于植物系统发育的构建和更深层次的理解物种的演化历程。

参考文献：

DONOGHUE PC, YANG Z, 2016. The evolution of methods for establishing evolutionary timescales[J]. Phil Trans Roy Soc B: Biol Sci, 371(1699).

HUG LA, ROGER AJ, 2007. The impact of fossils and taxon sampling on ancient molecular dating analyses[J]. Mol Biol Evol, 24(8): 1889-1897.

JIAO Y, WICKETT NJ, AYYAMPALAYAM S, et al., 2011. Ancestral polyploidy in seed plants and angiosperms[J]. Nat, 473(7345): 97-100.

KRESS WJ, SOLTIS DE, KERSEY PJ, et al., 2022. Green plant genomes: What we know in an era of rapidly expanding opportunities[J]. Proc Nat Acad Sci, 119(4).

LANFEAR R, HO SYW, JONATHAN DAVIES T, et al., 2013. Taller plants have lower rates of molecular evolution[J]. Nat Comm, 4(1): 1879.

LANFEAR R, WELCH JJ, BROMHAM L, 2010. Watching the clock: studying variation in rates of molecular evolution between species[J]. Trend Ecol Evol, 25(9): 495-503.

LI HT, YI TS, GAO LM, et al., 2019. Origin of angiosperms and the puzzle of the Jurassic gap[J]. Nat Plant, 5(5): 461-470.

LI L, STOECKERT CJ, ROOS DS, 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes[J]. Genome Res, 13(9): 2178-2189.

LUO A, DUCHÊNE DA, ZHANG C, et al., 2020. A simulation-based evaluation of tip-dating under the fossilized birth–death process[J]. Syst Biol, 69(2): 325-344.

LUO J, ZHANG YP, 2000. Molecular clock and its existing problems[J]. Acta Anthropol Sinica, 19(2): 151-159. [罗静，张亚平，2000. 分子钟及其存在的问题[J]. 人类学学报，19(2): 151-159.]

MAGALLÓN S, 2010. Using fossils to break long branches in molecular dating: a comparison of relaxed clocks applied to the origin of angiosperms[J]. Syst Biol, 59(4): 384-399.

REN R, WANG H, GUO C, et al., 2018. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms[J]. Mol Plant, 11(3): 414-428.

SHANG J, TIAN J, CHENG H, et al., 2020. The chromosome-level wintersweet (*Chimonanthus praecox*) genome provides insights into floral scent biosynthesis and flowering in winter[J]. Genome Biol, 21(1): 1-28.

SILVESTRO D, BACON CD, DING W, et al., 2021. Fossil data support a pre-Cretaceous origin of flowering plants[J]. Nat Ecol Evol, 5(4): 449-457.

SMITH SA, DONOGHUE MJ, 2008. Rates of molecular evolution are linked to life history in flowering plants[J]. Sci, 322(5898): 86-89.

SONG X, SUN P, YUAN J, et al., 2021. The celery genome sequence reveals sequential paleo-polyploidizations, karyotype evolution and resistance gene reduction in apiales[J]. Plant Biotechnol J, 19(4): 731-744.

SONG X, WANG J, LI N, et al., 2020. Deciphering the high-quality genome sequence of coriander that causes controversial feelings[J]. Plant Biotechnol J, 18(6): 1444-1456.

SUN P, JIAO B, YANG Y, et al., 2021. WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes[J]. BioRxiv. https://doi.org/10.1101/2021.04.29.441969.

TANG H, WANG X, BOWERS JE, et al., 2008. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps[J]. Genome Res, 18(12): 1944-1954.

TANG XH, LAI XL, ZHONG Y, 2002. Molecular clock hypothesis and fossil record[J]. Earth Sci Front, 9(2):465-474. [唐先华，赖旭龙，钟扬，等，2002. 分子钟假说与化石记录[J]. 地

学前缘，9(2)：465-474.]

VANNESTE K, VAN DE PEER Y, MAERE S, 2013. Inference of genome duplications from age distributions revisited[J]. Mol Biol Evol, 30(1): 177-190.

WANG J, SUN P, LI Y, et al., 2018. An overlooked paleotetraploidization in Cucurbitaceae[J]. Mol Biol Evol, 35(1): 16-26.

WANG J, SUN P, LI Y, et al., 2017. Hierarchically aligning 10 legume genomes establishes a family-level genomics platform[J]. Plant Physiol, 174(1): 284-300.

WANG J, YUAN J, YU J, et al., 2019. Recursive Paleohexaploidization shaped the durian genome[J]. Plant physiol, 179(1): 209-219.

WANG S, XIAO Y, ZHOU ZW, et al., 2021. High-quality reference genome sequences of two coconut cultivars provide insights into evolution of monocot chromosomes and differentiation of fiber content and plant height[J]. Genome Biol, 22(1): 1-25.

WANG X, GUO H, WANG J, et al., 2016. Comparative genomic de-convolution of the cotton genome revealed a decaploid ancestor and widespread chromosomal fractionation[J]. New Phytol, 209(3): 1252-1263.

WANG X, WANG J, JIN D, et al., 2015. Genome alignment spanning major Poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events[J]. Mol Plant, 8(6): 885-898.

WU S, HAN B, JIAO Y, 2020. Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms[J]. Mol Plant, 13(1): 59-71.

YANG Y, SUN P, LV L, et al., 2020. Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution[J]. Nat Plant, 6(3): 215-222.

YANG Z, 2007. PAML 4: phylogenetic analysis by maximum likelihood[J]. Mol Biol Evol, 24(8): 1586-1591.

ZHANG L, CHEN F, ZHANG X, et al., 2020a. The water lily genome and the early evolution of flowering plants[J]. Nat, 577(7788): 79-84.

ZHANG L, WU S, CHANG X, et al., 2020b. The ancient wave of polyploidization events in flowering plants and their facilitated adaptation to environmental stress[J]. Plant Cell Environ, 43(12): 2847-2856.

ZHUANG W, CHEN H, YANG M, et al., 2019. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication[J]. Nat Genet, 51(5): 865-876.